A DEEP LEARNING FRAMEWORK FOR DETECTION AND SEGMENTATION OF MULTIPLE ARTEFACTS IN ENDOSCOPIC IMAGES

1. INTRODUCTION

Endoscopy is a standard diagnostic and therapeutic procedure for examining hollow organs. The endoscopic procedure and tool name vary depending on the organ under examination. Colonoscopy for analysis of the colon and gastroscopy to inspect the upper gastrointestinal tract etc. [1]. In every procedure, the clinician views the organ using a monitor connected outside. The frames displayed are affected by artificial effects called artefacts. They are not present in the organ in reality but occur due to variations in handling the scope like hand movements, lighting conditions, modality of imaging and natural phenomena. They affect the visual interpretation of clinicians, increase false detection rates of automated diagnosis systems and challenge the robustness of Artificial Intelligence (AI)-based methods deployed in Computer-Aided Diagnostic (CAD) systems. Artefacts are common in every modality of medical imaging [2][3]. Thus, as an image interpreter, it is vital to be aware of artefacts and their background. This research addresses artefacts that occur predominantly in endoscopic images of organs like the esophagus, stomach, duodenum, liver and colon. Various artefacts are present in endoscopic images, including saturation, contrast, specular reflections, motion blur, defocus blur, bubbles, instruments, blood, and miscellaneous artefacts like chromatic aberration, floating debris, etc. Prior knowledge of such artefacts avoids erroneous reporting. In the present era of digital medicine, thousands of such images are generated daily. There is a great demand for smart/ intelligent equipment to manage the situation. AI plays a significant role in medical imaging, from classification and disease detection to segmentation. Such technologies aid doctors towards diagnosis in a shorter time with improved accuracy.

Thus, joining hands with the state-of-the-art deep learning framework to detect and segment these artefacts could eventually bring out better endoscopic imaging systems. Minimal research is found in this area primarily due to a lack of well-annotated datasets. The available research outcomes mainly focus on improving the accuracy of artefact detection and segmentation system. But inference time is another crucial parameter when

incorporated into guided robotic surgery and CAD system. Thus, this research focuses on developing a well-annotated dataset, a state-of-the-art detection and segmentation model with improved accuracy and good inference time.

2. LITERATURE SURVEY

With the rapid developments in deep learning for computer vision, object detection and image segmentation algorithms have revealed rapid growth in recent years. It is due to the hidden feature representations provided by Convolutional Neural Networks (CNN) that show significant development over hand-crafted features. Amidst technology development to date, the missing clinical abnormality during the procedure is found to be more than 15% [4]. Therefore, the need for designing a robust technology increases day by day.

2.1 Detection and Localization of Endoscopic Artefacts

Object detection refers to identifying and localizing an object of interest from the image. This approach can be broadly classified into four main categories: single-stage, two-stage, multi-stage and anchor-free detectors.

a. Endoscopic artefact detection using single-stage object detectors

Single-stage detectors like You Only Look Once (YOLO), Single Shot Detector (SSD) and RetinaNet perform object detection in a single pass on the input image. Sharib Ali et al. [5] trained variants of YOLOv3 and proposed YOLOv3-Spatial Pyramid Pooling (SPP). It outperformed with best results in terms of mean Average Precision(mAP) and inference time. Seiryo Watanabe et al. [6] proposed a new scheme combining Mask (Region-based CNN) R-CNN and YOLOv3 for detecting endoscopic artefacts. Mask R-CNN was preferred for artefacts with clear and defined boundaries like specular reflections, instruments etc. and YOLOv3 for the blur and contrast.

Suyog Jadhav et al. [7] adopted RetinaNet with ResNet50 and Feature Pyramid Network (FPN) as the backbone of one of the models for ensembled architecture. Xiahong Gao and Barbara Braden [8] developed an architecture based on RetinaNet for detection and instance segmentation. Anand Subramanian and Koushik Srinivasan [9] adopted an ensembled model combining the RetinaNet network with ResNet50 and ResNet101 feature extractors. Vishnusai Y. et al. [10] trained RetinaNet architecture to extract baseline results.

b. Endoscopic artefact detection using two-stage object detectors

Girshick et al. [11] proposed an R-CNN network for object detection, the first of its kind to use a CNN in an object detection network for higher performance. He proposed a fast version of R-CNN called Fast R-CNN [12] in 2015. Later, an updated version called Faster R-CNN [13] was developed. Mask R-CNN [14], an extension of Faster R-CNN with a parallel branch, for instance segmentation, was developed in 2018.

Xiaokang Wang and Chunqing Wang [15] took a Faster R-CNN model and applied the right training strategy to improve the network performance. Pengyi Zhang et al. [16] proposed a Mask aided R-CNN. The authors initially trained the original Mask R-CNN with a segmentation dataset. Later, the trained model was allowed to predict classes from the detection set.

c. Endoscopic artefact detection using a multi-stage object detector

Multi-stage detectors aim at achieving better accuracy than one-stage and two-stage detectors. Cascaded R-CNN [17], an extended version of Faster R-CNN, is said to overcome the problem of overfitting during training and inference time mismatch between the Intersection over Union (IoU). Qingtian Ning et al. [18] used basic cascaded R-CNN with L1 loss function for artefact detection. Yan-Yi Zhang and Di Xie [19] proposed a multi-stage cascaded R-CNN combined with FPN. The author used pre-trained weights and a phased approach to gradually increase the IoU threshold during training.

Suhui Yang and Guanju Cheng [20] improved the cascaded R-CNN structure by adding ResNet101 as the backbone and the FPN module. Gorkem Polat et al. [21] proposed an ensemble model in which cascaded R-CNN was one among the three models used. The prediction of all three models, namely Faster R-CNN, Cascaded R-CNN and RetinaNet, were ensemble for best performance.

d. Anchor-free detectors

Traditional deep learning-based architectures heavily rely upon anchors for predicting semantic objects in an image. These anchors have various scales and aspect ratios based on the object detected in an image. The speed and accuracy of the detector are based on the anchors, the fewer the anchors faster the detectors, but it may reduce accuracy. Detection of endoscopic artefacts using anchor-free object detectors is at its nascent stage.

2.2 Image Segmentation

Deep learning-based segmentation algorithms have an enormous impact on medical science. Long et al. [22] proposed a Fully Convolutional Network (FCN) for semantic segmentation. It became the first-ever high-impact model for convolutional neural network-based segmentation methods, where pixel-wise output is calculated by up-sampling the output activation maps. SegNet[23], the next notable architecture, was developed in 2015. The network up samples input feature maps of low resolution too. The same year, Ronneberg et al. presented a network architecture called U-Net [24], which addresses the vanishing gradient problem. The architecture was well-suited for biomedical image segmentation. Many authors suggested several modifications with the above network structures as a base.

a. Segmentation of endoscopic artefacts

All the proposed algorithms for endoscopic artefact segmentation show variation across segmenting different classes of artefacts and the difficulty in handling class overlap. It arises due to the irregular-shaped structures of the artefacts, which convolutional kernels find challenging to identify.

Suhui Yang and Guanju Cheng [20] took the Deeplabv3+ network with two different backbones, ResNet101 and MobileNet, for segmentation. They proved that merging the backbones improved the algorithm's performance. Yun Bo Guo et al. [25] took DeepLab v3+ architecture as a base for semantic segmentation and proposed a novel architecture called DeepEAD. Shufan Yang and Sandy Cochran [26] have proposed an architecture based on U-Net. ResNet50, pre-trained on ImageNet, was used in the architecture as a backbone for extricating features for semantic segmentation.

Hoang Manh Hung et al. [27] reformed U-Net with a SE-ResNeXt50 backbone. The author selected the network as it tackles the class imbalance problem and can retrieve the hidden fragments in the images. Anand Subramanian and Koushik Srivatsan [9] considered two models: U-Net with ResNet50 backbone and DeeplabV3 with Xception backbone. The team used weights pre-trained on the ImageNet dataset. Le Duy Huýnh and Nicolas Bountry [28] proposed different architectures based on U-Net and U-Net++. EfficientNetB1 was used as the backbone, which was expected to have a balance between resolution, network depth and width.

2.3 Summary

Research works in the literature use either a single-stage object detector or a two-stage object detector for artefact detection. The researchers experimented with different backbone networks, modified existing architectures, performed cascading of current models, and more. After a detailed review, a mAP of 47.91% has been achieved to detect endoscopic artefacts. Similarly, for artefact segmentation, most researchers notably used existing state-of-the-art algorithms with various backbones, employed an ensemble approach, and introduced different augmentation techniques. The researchers used the above methods to capture the most delicate features of every artefact and segment the same to the best. The segmentation performance in terms of Jaccard and F2 scores has reached 0.753 and 0.796, respectively. Yet the number of studies is found to be limited. It opens an arena for researchers to bring out promising results.

3. PROBLEM FORMULATION

The primary challenges identified behind implementing such AI-powered artefact detection and segmentation system are as follows:

- The tissue appears different in different modalities of imaging.
- Artefacts overlap each other, leading to over-estimation of the bounding box.
- Different imaging artefacts occur due to various physical phenomena.
- Frequent occurrence of artefacts in clusters in every frame.
- The size, location, and nature of artefacts vary, lacking prominent geometrical structures.
- The Endoscopic Artefact Detection (EAD) dataset is the only source of annotated datasets.

Amidst the challenges mentioned above, a clinician or a CAD system must diagnose diseases, which is time-consuming and may lead to erroneous results.

4. OBJECTIVES OF THE RESEARCH

The following are the objectives set for the research:

- To collect endoscopic images from hospitals comprising Indian patients and create a new dataset.
- To annotate artefacts like specularity, saturation, blur, contrast, bubbles, instruments, blood and miscellaneous artefacts.

- To train various existing deep learning architectures for artefact detection on public and custom datasets, tune hyperparameters and set benchmarks.
- To design an ensembled architecture for endoscopic artefact detection for outstanding results.
- To design and train deep learning architecture for artefact segmentation to segment five different artefacts, namely saturation, miscellaneous artefacts, bubbles, instruments and specular reflections and produce outstanding results.

5. PROPOSED METHODOLOGY

A step-by-step methodology of the research plan is given below:

Step 1: Collection of datasets.

Step 2: Training of artefact detection and segmentation algorithm.

Step 3: Testing of the trained model with images from the test set.

Step 4: Evaluation of trained algorithms.

Step 5: Fine-tune hyperparameters and retrain the algorithm based on evaluation Scores.

5.1 Dataset

The lifeline of many real-time AI-based applications is in the hands of well-annotated datasets and training hardware. Both play a dominant role in the field. Deep Learning algorithms learn by themselves from the labelled data. Thus, a well-annotated dataset is essential.

a. Public Dataset

Many datasets are publicly available for research purposes. The available dataset covers only a single artefact for a specific application, like artefact detection in endoscopic images. So, researchers in the last decade concentrated on detecting a single artefact. The EAD dataset came into existence in the year 2019. It motivated many researchers to take the challenge of detecting multiple artefacts present in the image in a single shot.

EAD2019 [29] is a dataset that covers seven major imaging artefacts like specularity, saturation, contrast, blur, bubbles, instruments and miscellaneous artefacts. The dataset contains 2147 images. EAD2020[30] covers eight imaging artefacts which include blood, along with all seven artefacts covered by the EAD2019 dataset. It comprises

2531 images. Both the datasets are multi-patient, i.e., the images were collected from 6 different centres worldwide. It is a multi-organ, multi-tissue and multi-modality dataset. The images hosted in the dataset do not contain any patient information. The dataset contains images and a binary mask for semantic segmentation.

b. Custom Dataset

The public dataset contains images of patients from western countries. The images with artefacts like blur, instrument and saturation are not much found. Therefore 2400 endoscopic images comprising the Indian patients are collected. More images with the artefact saturation, blur and instrument to balance the dataset are collected. All the images are analysed to list out possible artefacts and decided to annotate for the following artefacts: specularity, saturation, blur, bubbles, blood, contrast, instruments and miscellaneous imaging artefacts. Initially, images are annotated in the presence of a senior clinician to gain expertise. Later, annotations are done by the scholar and finally validated by the clinician in the ratio of 1:10(no. of images).

c. Annotation protocols

Annotation protocols of the EAD dataset are followed for annotating the custom dataset. The endoscopic artefact detector is also trained using images from EAD datasets. Uniform annotation protocols from EAD datasets are used to maintain homogeneity across all three datasets.

d. Annotation software

All annotations are done using VGG (Visual Geometry Group) Image Annotator (VIA), an open-source annotation tool offered by Oxford University, United Kingdom.

e. Training and Test Split

All three datasets are subdivided as follows. The dataset is first split into 80% of images for training and 20% for testing. Later the training set is again divided into two training and validation sets. Thus, on the whole, 70% of images are for training, 10% for validation and 20% for testing.

f. Data Augmentation

Data augmentation is used to synthetically expand the existing dataset to suit the deep learning model data requirement. In data augmentation, some alterations will be made

to the original image to create a transformed version of the original images. When such techniques are deployed, the detector can be trained and made more robust. Standard data augmentation techniques include position augmentation and colour augmentation. The former embraces rotation, padding, scaling, cropping, flipping, translation and affine transformation. The latter method entails varying hue, saturation, brightness and contrast. The selection of augmentation techniques depends upon the application. The performance of the detector varies based on the augmentation technique chosen.

Our research chooses manual data augmentation techniques like rotation and flipping. Rotation angles like 90°,180° and 270° are chosen for rotation, and the same images are flipped after rotation. Thus, after data augmentation, the dataset is multiplied by eight times the original dataset size.

g. Statement of Ethics

The collection of endoscopic images from hospitals requires Institute Human Ethical Clearance (IHEC). The process for the same is done, and clearance has been obtained from the IHEC Team of Avinashilingam Institute.

5.2 Endoscopic Artefact Detector

a. Methodology

This research uses three well-known object detection algorithms to detect endoscopic imaging artefact, namely YOLOv3, YOLOv4 and Faster R-CNN. YOLOv3 uses Darknet-53 as the backbone. The algorithm is said to have improved performance over other top-performing models. YOLOv4 is an upgraded version of YOLOv3. It uses Cross Stage Partial (CSP) darknet53 as the backbone, and both are single-stage object detectors. Both work towards faster predictions. Faster R-CNN is a two-stage object detector which is known for accuracy. All the three models considered are trained and ensembled for final predictions.

The training procedure of all the three state-of-the-art object detection models, namely YOLOv3, YOLOV4 and Faster R-CNN, is discussed in the following sections, along with the proposed ensemble model architecture. The datasets are initially divided into train, validation and test. The training set of images are data augmented. The augmented dataset is used to train all three models. The images in the train set are

maintained uniform across all three models. Intensive training is carried out using a high-performance Graphics Processing Unit (GPU) from Google Co-laboratory.

b. Training YOLOv3 for Endoscopic Artefact Detection

YOLOv3 uses Darknet53 as the backbone to extract features. It finds a good balance between the detection of various sizes of objects, from tiny to large. This characteristic is beneficial in the case of detecting artefacts like specular reflections, which are small and artefacts like contrast and instrument, which are prominent. Darknet is an open-source neural network. It can be cloned from the repository offered by the authors. It holds pre-trained weights and configuration files for various models of YOLO. The pre-trained weights and the configuration file are cloned from the author repository to attain the benefits of transfer learning. The training parameters are set up along with data augmentation initializations. The total number of classes to be detected is eight; thus, classes are set to eight in all three YOLO layers in the configuration file.

By setting all the necessary initializations, the training started. The training lasts until the network reaches a minimum average loss. Approximately after 55,000 iterations, the average loss curve flattens. Once the average loss no longer reduces, the iterations can stop. Hence training stopped at 70,000 iterations. Various weight files extracted during training are tested for their performance concerning mAP and IoU. Scores obtained at the $65{,}000^{th}$ iteration are found the best.

c. Training YOLOv4 for Endoscopic Artefact Detection

YOLOv4 offers normalization, data augmentation, and regularization techniques to improve model accuracy without compromising the model inference time. On the other hand, it is packed with a spatial attention module, non-linear activation, and Non-Max Suppression (NMS) to improve accuracy at the cost of inference time. Thus, the researcher must select the best techniques for the best results. For this research, YOLOv4 with CSPdarknet53 as the backbone, PanNet for aggregating the features and YOLOv3 head for final predictions are chosen. Special features of YOLOv4 like mosaic augmentation, Mish Activation function, NMS, optimized anchors etc., are used. Pre-trained weights for YOLOv4 are used to avoid training from scratch. The training lasts 95,000 iterations. Weights files extracted during the training process are examined for best results. Weight file extracted at $76{,}000^{th}$ iteration gives a good balance between mAP, IoU and inference time.

d. Training Faster R-CNN for Endoscopic Artefact Detection

For this research, Faster R-CNN is adopted from detectron2. It is built by Facebook AI Research (FAIR). It supports the rapid implementation of many object detection algorithms like mask R-CNN, faster R-CNN, RetinaNet etc. It also supports instance image segmentation and panoptic segmentation. It holds a model zoo consisting of trained model files for faster implementations and sets several baselines for research. With all the customized input parameters set, the training of Faster R-CNN started. Batch-wise, the training lasted until the model reached 80,000 iterations, out of which 74,000th iteration results surpassed all.

e. Ensemble model

The term 'ensemble' means collective or collaborative. Ensemble learning model combines the predictions of multiple object detection models to improve the overall performance. It combines the benefits of both single-stage and two-stage object detectors. The trained and tested model files of YOLOv3, YOLOv4 and Faster R-CNN are merged for predictions. A test image is passed into the model. All three trained models predict every artefact present in the image. Based on the ensemble method chosen, final predictions will be generated. In this research, a consensus approach is chosen for the ensemble where the final model accepts the predictions only if most models predict the same. The flow chart exhibited in figure 1 demonstrates the artefact detection process. The proposed research aims to develop an artefact detection model, which could be deployed as a part of the endoscopic imaging pipeline. Three different artefact detectors are combined to form an ensemble model, which has the advantages of single and two-stage detectors. The ensemble model's base learners are trained using EAD and Custom dataset's images. This proposed model will effectively detect all eight significant artefacts found in an endoscopic image.

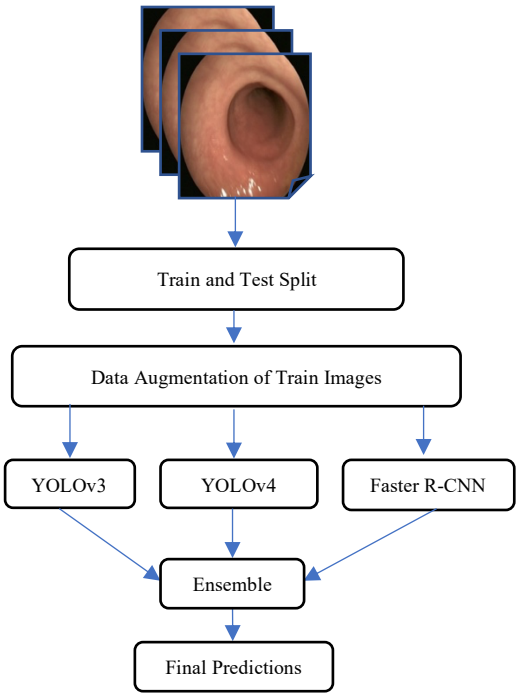


Figure 1 Ensemble Artefact Detection Model

5.3 Endoscopic Artefact Segmentation

The proposed artefacts segmentation model combines three deep learning models, namely variants of U-Net and LinkNet. All three models are trained with images from the EAD segmentation dataset. The dataset holds 675 images for segmentation of five major artefacts: saturation, specular reflections, bubbles, instruments and miscellaneous artefacts. Uniform training parameters were set for all three models. All three deep learning models are cloned from segmentation models Application Programming Interface (API) [31]. Segmentation models hold four different deep neural network architectures for image segmentation and 25 different backbones for each deep neural network. With all the parameters set, the training lasts till 300 epochs. The trained model is tested with images from the test set. The predictions and performance of the model are evaluated using standard evaluation metrics described in section 6. After the performance evaluation of the trained models, all three models are combined. Input images from the test set are passed through the proposed model. Each of the three models predicts the binary mask for the artefacts present in the images, and the predictions are averaged to form the final binary mask. Otsu method is used to convert the average mask into a binary mask. The final predictions of the proposed model are re-evaluated with the standard metrics. The result outperformed all existing models. The flow diagram displayed in figure 2 outlines the

segmentation process. The predictions from all three models and the final predictions are all extracted in binary format and evaluated against the ground truth binary mask.

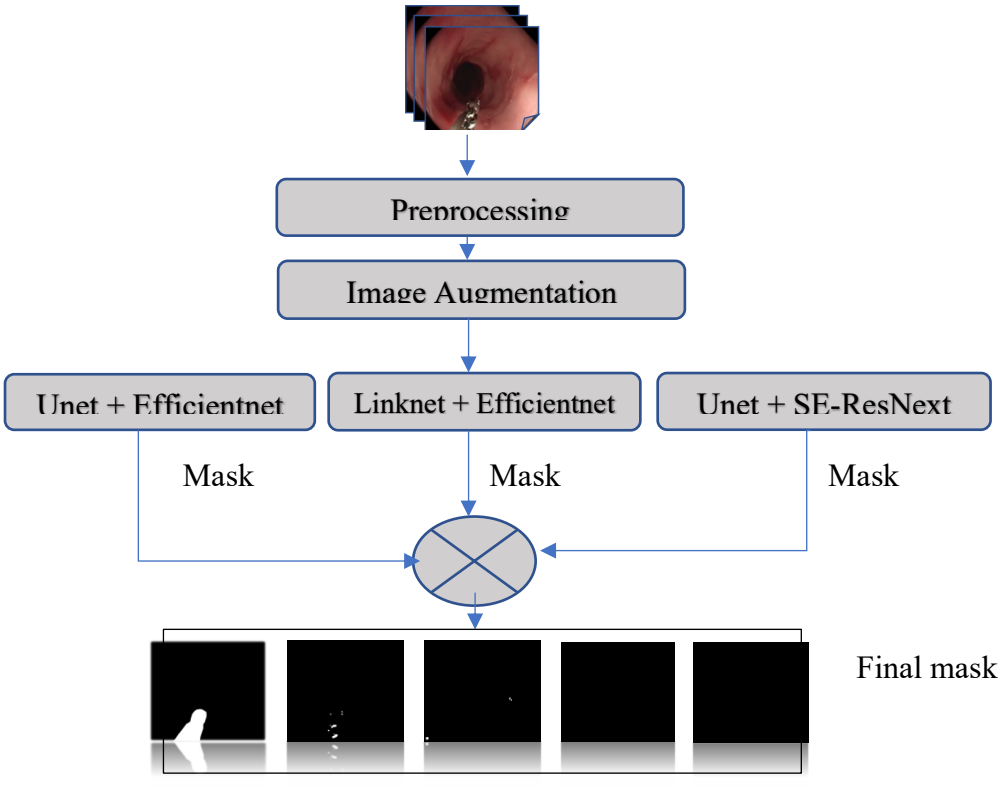


Figure 2 Proposed Artefact Segmentation Model

6. RESULTS AND DISCUSSION

6.1 Evaluation Criteria

Deep learning models are evaluated using images from the validation/test set. Performance metrics for assessing an object detection algorithm include precision, recall, mAP and IoU. The choice of metric depends on the nature of the application.

a. Intersection over Union (IoU)

IoU can be calculated using the formula in (6.1). IoU is a ratio between the intersection of ground truth(A) and predicted bounding boxes(B) and the union of ground truth(A) and predicted bounding boxes(B).

$$IoU = \frac{A \cup B}{A \cap B} \tag{6.1}$$

IoU is used to get true and false positives. A threshold value of 0.5 is set for classifying the same.

b. Mean Average Precision

The next metric to be calculated is the Average Precision (AP) which is obtained by intersecting the coordinate axis and the precision-recall curve at recall values (r1,r2,…). Equation (6.2) is used to calculate the AP score.

$$AP = \sum_i (r_{n+1} - r_n) p_{interp}(r_{n+1}) \tag{6.2}$$

where pinterp= max p(r). Finally, mAP is the mean of all average precisions calculated using the formula in (6.3).

$$mAP = \frac{1}{N} \sum_i AP_i \tag{6.3}$$

In this research, we have validated the coherent results of an ensemble object detection model using the performance metrics aforesaid.

c. Dice Coefficient or Dice Similarity Coefficient (DSC) or F1 Score

It can be calculated using the formula in (6.4)

$$Dice\ Coefficient\ or\ F1\ Score = \frac{2*|A \cap B|}{|A|+|B|} \tag{6.4}$$

In the above expression, A and B are two images, $|A \cap B|$ denotes pixels that are common to both the images, $|A| + |B|$ represents the sum of all the pixels in A and B. The score varies between 0 and 1. A score of 0 shows no similarity, and a score of 1 signifies a higher similarity between the predictions and ground truth image.

d. Generalized formula for Fß Score:

The generalized formula for calculating F1 and F2 scores is given in (6.5)

$$F_\beta\ Score = (1 + \beta^2) \frac{Precision*recall}{(\beta^2 * Precision)+recall} \tag{6.5}$$

By assigning ß=1 and ß=2, we can compute F1 and F2 scores. In the expression, the metric called precision is used to calculate the percentage of correct predictions of positive classes and recall is used to calculate the percentage of accurate predictions of positive classes out of all positive predictions made by the network. Precision and recall can be calculated using the formula shown in (6.6) & (6.7).

$$Precision = \frac{TP}{TP + FP} \tag{6.6}$$

$$Recall = \frac{TP}{TP + FN} \tag{6.7}$$

where TP represents True Positive (positive samples predicted as positive samples), FP represents False Positive (negative samples predicted as positive samples), and FN stands for false negative (positive samples predicted as negative samples).

e. $Score_s$

It is a weighted score calculated using common metrics. The formula is given in (6.8)

$$Score_S = 0.75 * [\,0.5 * (DSC + IoU)] + 0.25 * F_2Score \quad (6.8)$$

6.2 Performance of Endoscopic Artefact Detector

All the researchers used the EAD dataset for their research. Few used various augmentation techniques to enhance data set size, which include flipping, rotation, manual cropping, extracting patches etc. Some of the researchers used test time augmentation. Similarly, many researchers reported that the dataset suffers from a class imbalance problem. They approached the research problem from various angles, right from choosing the data augmentation technique, Object detection model, employing transfer learning, cascading existing architecture, an ensemble of various proposed models, modifying existing architecture etc.

On the other hand, the proposed model is trained with images from EAD and a custom dataset. The custom dataset is curated to add more images with blur, saturation and instruments to balance the class distribution and add images of Indian patients. The models used for the ensemble are trained with all the datasets. Figure 3 self-explains the predictions made by the ensemble model. It is found that the accuracy in identifying every artefact is more than 80%.

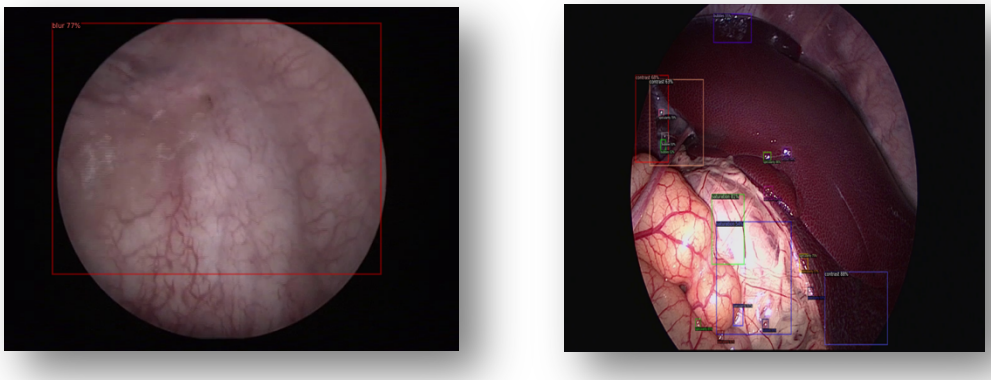


Figure 3 Predictions of Ensemble Model

The research results of existing models are compared with the proposed work through graphical representation in figure 4. The result is compared using standard metrics like mAP and IoU. The proposed method outperforms the existing results.

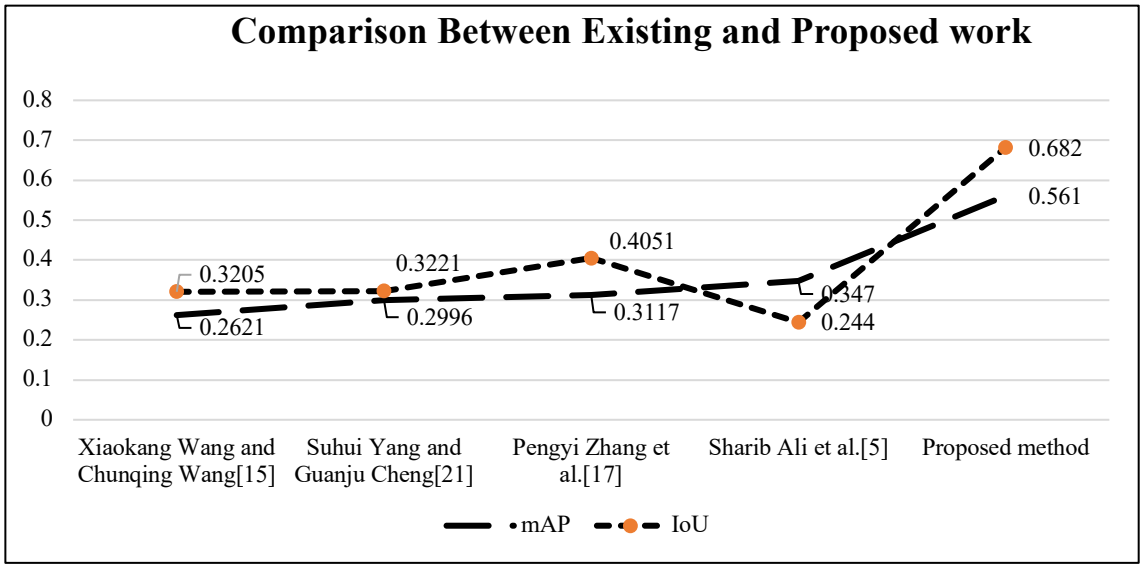


Figure 4 Comparison of Research Outcomes of Existing and Proposed Work for Artefact Detection

The other primary objective of the research is inference time, an important parameter in real-time applications. The proposed model reached 80.4ms compared to 88ms of the existing model, which is portrayed graphically in figure 5. Thus, the results show that the proposed model performs well in terms of mAP, IoU and inference time.

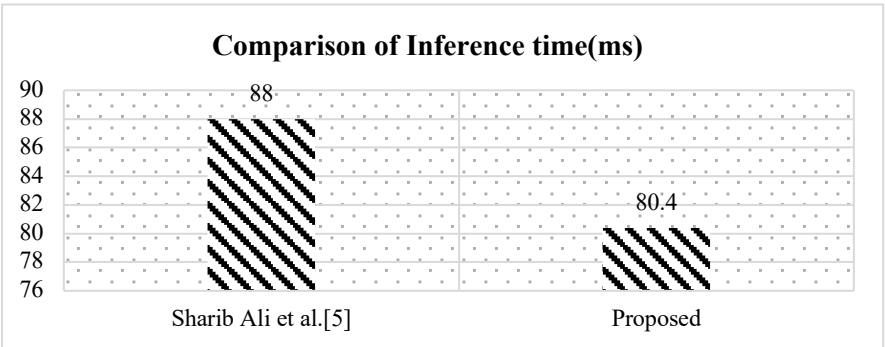
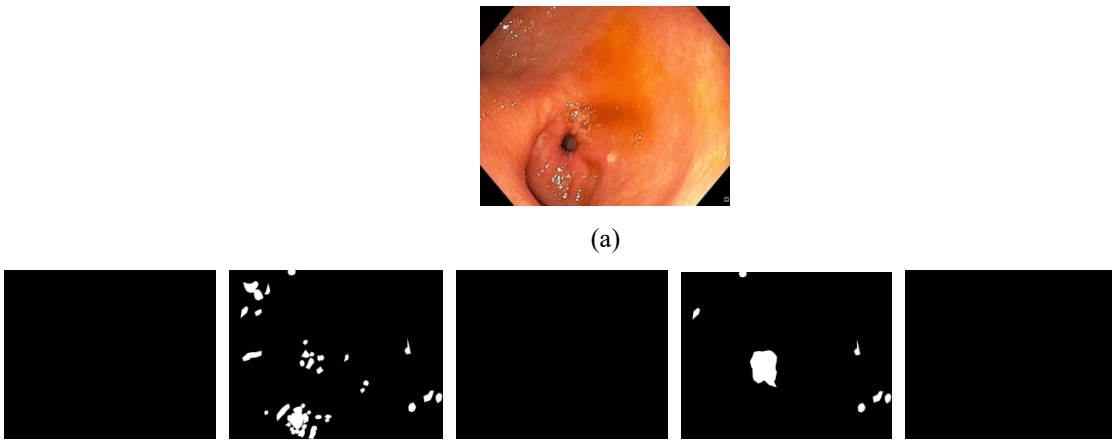


Figure 5 Comparison of Inference time

6.3 Performance of Endoscopic Artefact Segmentation

The algorithm's performance is assessed against existing results using standard performance evaluation metrics like dice score and F2 score. The proposed model incorporates

three learners whose predictions are averaged for final predictions. All three base learners are trained equally with the same train set and standard input parameters. The images in the dataset are found to be very limited. Therefore, data augmentation techniques are employed. The methods include affine transformation, scaling, random cropping, varying hue, saturation, brightness, rotation and flipping. After training for 300 epochs, the model performance is evaluated, tuned and retrained for best results. After finalizing the individual results, all three models are combined. Now the input image is passed on through the proposed model, which holds three learners. Each predicts the binary mask; all three masks are averaged for the final performance. The average mask is converted back to binary form with the help of the traditional Otsu method. Figure 6 shows the prediction of the final model extracted in binary mask format for each endoscopic imaging artefact under study.



(a) (b) (c) (d) (e) (f)

Figure 6 (a)-(f) Proposed segmentation results

(a)-Input image (c) Specular reflection segmentation (e) Bubbles segmentation results

(b),(d),(f)- Empty masks as no other artefacts are found

The results presented in figure 7 are obtained from the above-said model and compared with results of existing works. It is apparent from the graphical representation that the proposed model performs relatively better than previous methods.

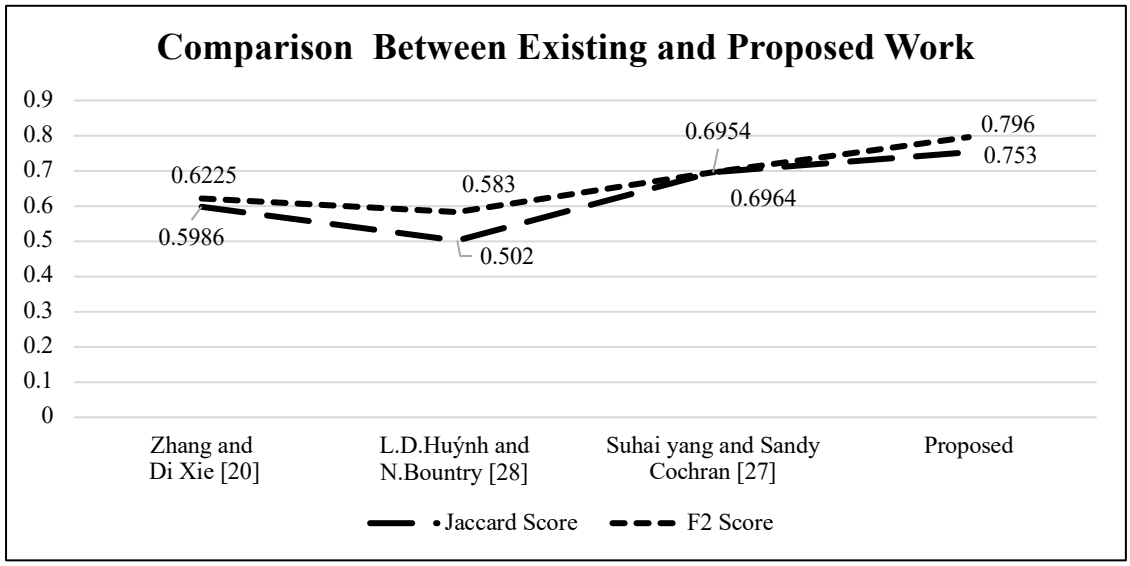


Figure 7 Comparison of Research Outcomes Between Existing and Proposed Work for Artefact Segmentation

6.4 Summary

In a nutshell, the various object detection models trained for the ensemble model produced the best results. The individual models are trained, tested and tuned to give the best performance and later combined for the ensemble. The consensus method of the ensemble provides the best results among the other two models. The proposed model performs best in terms of mAP, IoU and inference time. Similarly, the proposed segmentation model combines three trained deep neural network architectures whose individual performance is tuned towards the best results and later combined for predicting the final averaged mask. The evaluation results of the final model are found to be outstanding.

7. ORGANIZATION OF RESEARCH

The thesis is organized as follows:

Chapter 1 describes endoscopy, type of endoscopy, endoscopic imaging pipeline, and artefacts in endoscopic images with sample images. **Chapter 2** reviews the literature related to this research work. **Chapter 3** discusses the collection of datasets, training and test splitting and data augmentation techniques used for both artefact detection and segmentation. **Chapter 4** details the proposed ensemble artefact detector and details training the base learners. **Chapter 5** describes the proposed method for artefact segmentation and various traditional and deep learning algorithms trained during research.

Chapter 6 discusses a simple application, an endoscopic imaging pipeline for polyp classification. **Chapter 7** discuss the results obtained in detail. **Chapter 8** concludes the research work and presents the future scope of the research.

REFERENCES

[1]. Various types of endoscopy: Available at https://www.cancer.net/navigating-cancer-care/diagnosing-cancer/tests-and-procedures/types-endoscopy. Accessed on 13th April 2022.

[2]. Artefacts in medical images: https://radiopaedia.org/articles/radiological-image-artifact. Accessed on 18th May 2022.

[3]. Artefacts in medical images: https://www.upstate.edu/radiology/education/rsna/radiography/artifact.php. Accessed on 18th May 2022.

[4]. S. Ali et al., "Deep Learning for Detection and Segmentation of Artefact and Disease Instances in Gastrointestinal Endoscopy", Medical Image Analysis, Vol. 70,2021, https://doi.org/10.1016/j.media.2021.102002.

[5]. S. Ali, F. Zhou, A. Bailey, B.Braden, J.E. East, X. Lu, J. Rittscher, "A Deep Learning Framework for Quality Assessment and Restoration in Video Endoscopy", Medical Image Analysis, Vol. 68,2021, pp.1-25.

[6]. Seiryo Watanabe et al., "DNN Models and Post Processing Thresholds for Endoscopy Artefact Detection in Practice", Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366,paper-7.

[7]. Suyog Jadhav et al., "Multi-Plateau Ensemble for Endoscopic Artefact segmentation and Detection", Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 22-25.

[8]. X. Gao and B. Braden, "Artefact Detection and Segmentation Based on Deep Learning System", Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 80-81.

[9]. A. Subramanian, K. Srinivasan, "Exploring Deep Learning-based Approaches for Endoscopic Artefact Detection and Segmentation", Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp.51-56.

[10]. Vishnu Sai et al., " A submission Note on EAD2020: Deep Learning Based Approach for Detections Artefacts in Endoscopy", Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 84-85.

[11]. Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," pp. 1-21, 2014, arXiv: 1311.2524v5.

[12]. Ross Girshick, “Fast R-CNN,” Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015. DOI: 10.1109/ICCV.2015.169.

[13]. S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, Vol. 1,07-12 December 2015, pp. 91-99.

[14]. Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross Girshick, “Mask R-CNN,” pp.1-12, 2018, arXiv: 1703.06870v3.

[15]. X. Wang, C. Wang, “Detect Artefacts of various Sizes on the Right Scale for each Class in Video Endoscopy”, Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366.

[16]. P. Zhang, X.Li, Y. Zhong, “Ensemble Mask Aided R- CNN”, Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366.

[17]. Zhaowei Cai and Nuno Vasconcelos, “Cascade R-CNN: Delving into High-Quality Object Detection,” 2107, pp. 1-9, arXiv:1712.00726.

[18]. Qingtian Ning, xu Zhao, Jingyi wang(2019), “ Deep Layer Aggregation Approaches for Region Segmentation of Endoscopic Images”, Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366,paper-8.

[19]. Yan-Yi-Zhang and Di-Xie, “ Detection and Segmentation of Multi-class Artefacts in Endoscopy”, Journal of Zhejiang University, Science-B, vol. 20, issue 12, 2019, pp. 1014-1020.

[20]. Suhui Yang and Guanju Cheng (2019), “Endoscopic Artefact Detection and Segmentation with Deep Convolutional Neural Network”, Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366,paper-1.

[21]. Gorkem Polat, Deniz Sen, Alperen Inci and Alptekin Temizel, “Endoscopic Artefact Detection with Ensemble of Deep Neural Network and False Positive Elimination,” Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 8-12.

[22]. J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” arXiv:1411.4038, 2015, pp. 1–10.

[23]. V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, 2017, pp. 2481–2495.

[24]. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv:1505.04597, pp. 1–8, 2015.

[25]. Y. B. Guo, Q. Zheng, and B. J. Matuszewski, “Deep Encoder-Decoder Networks for Artefact Segmentation in Endoscopic Images,” Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 18-21.

[26]. S. Yang, and S. Cochran, "Graph-Search Based U-Net-D for the Analysis of Endoscopic Images," Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection: Multi-class Artefact Detection in Video Endoscopy, EAD2019, Venice, Italy, 8th April 2019, Vol. 2366, pp. 1-6.

[27]. H. M. Hung, P.T.D. Thinh, H.J Yang, S.H. Kim, G.S.Lee, "Artefact Detection and Segmentation using Cascaded R-CNN & U-Net", Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595, pp. 47-50.

[28]. L.D.Huýnh and N.Bountry, "A U-Net++ with Pre-trained EfficientNet Backbone for Segmentation of Diseases and Artefacts in Endoscopic Images and Videos," Proceedings of the 2nd International Workshop and Challenge on Computer Vision in Endoscopy, EndoCV 2020, Lowa City, United States, 3rd April 2020, Vol. 2595. Pp. 13-17.

[29]. Ali S. et al. "Endoscopy Artifact Detection (EAD 2019) Challenge Dataset,"pp-1-12, 2019. arXiv:1905.03209.

[30]. Ali, Sharib et al., "Endoscopy Artefact Detection (EAD) Dataset (includes updated 2020 version)", Mendeley Data, V3, 2021, DOI: 10.17632/c7fjbxcgj9.3.

[31]. Segmentation model API: https://segmentation-models.readthedocs.io/en/latest/api.html. Accessed on May 10, 2022.

PUBLICATION DETAILS

Journal Publications:

Sl.No.	Title of the Paper	Journal	Volume/Page. No./Month and Year of Publication
1.	Detecting Multi-Class Artifacts in Endoscopic Images using YOLOv3.	International Journal of Recent Technology and Engineering.	Vol. 9, No. 1, pp.-2578-2583, 2020. DOI:10.35940/ijrte.A3050.059120 **(Citations – 2)**
2.	Deep Learning Techniques to Detect Artefacts in Endoscopic Images: A Retrospective Review	Stochastic Modeling and Applications	Vol. 26, No. 3, pp. 94-104, 2022, ISSN: 0972-3641 **(UGC CARE I)**
3.	Transfer learning based Deep Neural Network for Detecting Artefacts in Endoscopic Images	International Journal of Electrical and Computer Engineering Systems	Vol. 13, No. 8, pp. 633-641, 2022, DOI: https://doi.org/10.32985/ijeces.13.8.3 ISSN:1847-6996 E-ISSN**:**1847-7003 **(SCOPUS INDEXED)**

4.	Performance Analysis of Traditional Algorithms and Deep Neural Networks for Segmentation of Multiple Artefacts in Endoscopic Images	Sadhana - Academy Proceedings in Engineering Sciences	WoS/Scopus indexed **Submitted**

Conference Presentation:

S. No.	Title of the Paper	Name of the Conference and Venue	Date
1.	Deep Learning Architectures for Multi-class Artefact Segmentation in Endoscopic Images: A Review	1st International Conference on Power of Digital Technologies in Societal Improvement, IETE CHENCON'21, IETE Chennai Centre,2021.	04.06.2021 and 05.06.2021
2.	YOLOv4 for Multi-class Artefact Detection in Endoscopic Images	2021 3rd IEEE International Conference on Signal Processing and Communication, Coimbatore, India, pp-73-77, 2021. **(Scopus Indexed)** **(Citations – 1)**	13.05.2021 and 14.05.2021
3.	Endoscopic Artefact Detection: Current Practices	National Conference on Emerging Trends in Wireless Communication, Signal Processing and Networking at Avinashilingam Institute, Coimbatore, 2020	21.02.2020 and 22.02.2020